



Universidade de Brasília
Departamento de Estatística

Função de distribuição generalizada de valor extremo transmutada.

Bianca Souza de Paiva

**Brasília
2016**

Bianca Souza de Paiva

**Função de distribuição generalizada de valor extremo
transmutada.**

Orientadora:

Profa. Dra. **Cira Etheowalda Guevara Otiniano**

Monografia apresentada ao Departamento de
Estatística da Universidade de Brasília para
obtenção do título de Bacharel em Estatís-
tica.

Brasília
2016

SUMÁRIO

RESUMO	5
1 Introdução	7
2 Conceitos Preliminares	9
2.1 TÉCNICA DE TRANSMUTAÇÃO	9
2.2 DISTRIBUIÇÃO DE VALORES EXTREMOS GENERALIZADA - GEV	10
3 Distribuição GEV transmutada	12
3.1 DEFINIÇÃO	12
3.2 MOMENTOS	14
3.3 MEDIDAS DE CONFIABILIDADE	18
3.4 QUANTIS	19
3.5 ESTATÍSTICAS DE ORDEM	21
4 Estimação e Simulação	24
4.1 ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA (MV)	24
4.2 ILUSTRAÇÕES NUMÉRICAS	26
5 Análise em Dados Reais	32
6 Considerações Finais	37
REFERÊNCIAS	39

RESUMO

A transmutação é uma técnica de generalização de funções de distribuições, a partir da composição de uma distribuição acumulada base. Neste trabalho utiliza-se a transmutação quadrática, proposta por Shaw (2009), para obter a distribuição de valor extremo generalizada transmutada. Foram geradas algumas medidas da nova distribuição como os momentos, quantis, funções de sobrevivência e risco e as estatísticas. Realizou-se simulações de dados via método da transformada inversa, os parâmetros da distribuição foram estimados por máxima verossimilhança e as estimativas testadas via simulação Monte Carlo. Além disso, o modelo foi utilizado para ajustar três conjuntos de dados reais, relacionados a índices financeiros.

Capítulo 1

Introdução

Muitos problemas estatísticos têm interesse em obter informação à respeito de acontecimentos raros (de baixa frequência), que podem, por sua vez, gerar resultados desastrosos. Assim, a distribuição generalizada de valores extremos surge como uma proposta de modelo capaz de fornecer tais informações, mesmo com a escassez de dados provenientes das caudas das distribuições.

A distribuição generalizada de valores extremos (popularmente conhecida por GEV, do inglês: *generalized extreme value*) consiste de uma família que reúne distribuições de valor extremo desenvolvidas a partir da chamada teoria dos valores extremos. Esta teoria é baseada no estudo da distribuição limite das estatísticas do máximo normalizado (ou do mínimo) de variáveis aleatórias independentes, que gera a chamada "distribuição de valor extremo".

Os principais desenvolvedores da teoria, Fisher e Tippet (1928), mostraram que uma distribuição de valor extremo pode ser de um dos três tipos: Gumbel, Fréchet e Weibull, que também são conhecidas como distribuição de valor extremo tipo I, II e III, respectivamente. Posteriormente, Jenkinson [1955] mostrou que as três distribuições extremas mencionadas podem ser escritas em uma única distribuição, a distribuição generalizada de valores extremos.

Historicamente, as primeiras aplicações da teoria se concentraram na área de hidrologia, buscando analisar a frequência de fluídos em fenômenos como precipitações e inundações, e também em modelagens de terremotos e emissões radioativas, dentre outros. Porém, com o desenvolver da teoria, a aplicabilidade tem se tornado cada vez mais ampla, algumas das principais áreas que recentemente utilizam essa teoria são atuária, para cálculo da

probabilidade da ruína por eventos extremos, e finanças, para modelar ativos que apresentam cauda pesada.

No presente trabalho estudamos uma generalização da distribuição GEV, denominada GEV transmutada. A transmutação de uma distribuição, é uma técnica desenvolvida por Shaw e Buckley (2009), e consiste na introdução de mais um parâmetro à uma dada distribuição base, com a finalidade de inserir assimetria ou curtose na mesma, isso se dá através de composição da função de distribuição acumulada, quando esta possui inversa. Apesar de ser uma técnica relativamente nova, a transmutação já foi aplicada à várias outras funções de distribuição.

Assim, neste trabalho são investigadas as principais medidas de probabilidade da distribuição GEV transmutada, bem como a estimação dos parâmetros do modelo. Para isso, o conteúdo do trabalho está composto por quatro capítulos. No capítulo 2 apresentam-se resultados preliminares sobre a técnica de transmutação e a distribuição GEV. O terceiro capítulo trata da distribuição GEV transmutada e calcula algumas medidas de probabilidade importantes. A parte de inferência dos parâmetros está disposta no quarto capítulo. E, por fim, o capítulo 5 apresenta uma aplicação da estimação do modelo para dados reais.

Capítulo 2

Conceitos Preliminares

2.1 TÉCNICA DE TRANSMUTAÇÃO

A técnica de transmutação foi criada por Willian T. Shaw e Ian R. C. Buckley em (2009), com o intuito de conseguir obter uma distribuição assimétrica a partir de uma simétrica. A técnica foi amplamente aceita e na literatura já são conhecidas as transmutações de diversas distribuições simétricas e não simétricas. Alguns exemplos de distribuições transmutadas e seus autores são dadas a seguir: a transmutação da log-logística foi dada por Aryal (2013); da Gumbel e Weibull por Aryal e Tsokos (2009) e (2011), respespectivamente; da Exponencial, Rayleigh e Lindley-geométrica por Merovci (2013) ; Weibull-geométrica por Merovci e Elbata (2014); Fréchet por Mahmoud e Mandouh (2014).

No próximo capítulo, obtemos a distribuição GEV transmutada que engloba as distribuições transmutadas da Weibull (em seu suporte negativo) , Gumbel e Fréchet. Para tanto, nesta seção apresentamos a forma geral de uma distribuição transmutada.

A base da transmutação está consolidada como uma função da inversa da função de distribuição acumulada, $G_R(u) = G(F^{-1}(u))$, em que F e G possuem o mesmo suporte, sendo $F^{-1}(u) = \min\{x \in \mathbb{R} : F(x) \geq u\}$, a transformada inversa de F , para $u \in [0, 1]$.

A função de $G_R(u)$ percorre o intervalo unitário $I = [0, 1]$, e sob os pressupostos adequados satisfazem $G_R(0) = 0$ e $G_R(1) = 1$. O mapa de transmutação quadrática é definido como: $G_R(u) = u + \lambda u(1 - u)$, $|\lambda| \leq 1$, $0 \leq u \leq 1$, do qual resulta que as funções de distribuição acumuladas satisfazem a relação:

$$F(x) = (1 + \lambda)G(x) - \lambda G(x)^2.$$

Quando $G(x)$, função de distribuição acumulada da distribuição base, é absolutamente contínua, a derivada gera a função de densidade de probabilidade dada por:

$$f(x) = (1 + \lambda) g(x) - \lambda 2g(x)G(x),$$

em que $f(x)$ e $g(x)$ são as funções de densidade de probabilidade associadas com as distribuições acumuladas $F(x)$ e $G(x)$, respectivamente. Observa-se que para $\lambda = 0$ temos a distribuição da variável aleatória de base. Outro ponto a se notar é: quando o valor do parâmetro λ é negativo a função de densidade $f(x)$ possui a mesma estrutura de uma mistura da distribuição base com sua *skew* correspondente, isso porque $(1 + \lambda)$ e $(-\lambda)$ resultariam em valores entre 0 e 1, assim como são os pesos de uma mistura de distribuições e o segundo elemento da mistura ($2g(x)G(x)$) corresponde a generalização da distribuição $g(x)$ conhecida por skew G , introduzida por O'Hagan and Leonard (1976).

2.2 DISTRIBUIÇÃO DE VALORES EXTREMOS GENERALIZADA - GEV

A distribuição generalizada de valores extremos, proposta por Jenkinson (1955), reúne as três distribuições de valor extremos desenvolvidas por Fisher e Tippet (1928), que são: Gumbel, Fréchet e Weibull. Tais distribuições foram desenvolvidas na teoria dos valores extremos e são obtidas através do limite de máximos (ou mínimos) normalizados da seguinte maneira:

$$\lim_{n \rightarrow \infty} P \left(\frac{M_n - b_n}{a_n} \leq x \right) = [F(a_n x + b_n)]^n \rightarrow G(x)$$

onde M_n representa $\max\{X_1, X_2, \dots, X_n\}$ (ou o mínimo), em que X_1, X_2, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas (iid), e $a_n > 0$ e b_n são constantes de normalização.

Desta forma, uma variável aleatória X é definida como generalizada de valores extremos com parâmetros de forma (índice caudal) γ , locação μ e escala $\sigma > 0$ se sua função de densidade de probabilidade (fdp) é dada por:

$$g(x) = \begin{cases} \frac{1}{\sigma} \left[1 + \gamma \frac{(x-\mu)}{\sigma} \right]^{-\frac{1}{\gamma}-1} e^{\left\{ -[1+\gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}}, & \gamma \neq 0 \\ \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}} e^{-e^{-\frac{(x-\mu)}{\sigma}}}, & \gamma = 0 \end{cases} \quad (2.1)$$

E tem função de distribuição acumulada (fda) dada por:

$$G(x) = \begin{cases} e^{\left\{ -[1+\gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}}, & \gamma \neq 0 \\ e^{-e^{-\frac{(x-\mu)}{\sigma}}}, & \gamma = 0 \end{cases} \quad (2.2)$$

O domínio da distribuição, quando $\gamma = 0$ é composto por toda a reta dos reais já quando $\gamma \neq 0$, o domínio é definido pela expressão $1 + \gamma \frac{(x-\mu)}{\sigma} \geq 0$. O parâmetro de forma

da distribuição GEV também define à qual das três distribuições de valor extremo o modelo se relaciona, sendo que para $\gamma = 0$ a GEV está associada à distribuição Gumbel, quando $\gamma > 0$ relaciona-se à Fréchet e quando $\gamma < 0$ segue Weibull (em seu suporte negativo).

Capítulo 3

Distribuição GEV transmutada

3.1 DEFINIÇÃO

A técnica de transmutação, proposta por Shaw e Buckley, no ano de 2009, consiste em uma generalização de distribuições, tornando-as mais flexíveis. Motivados, na área de matemática financeira, pela necessidade de famílias paramétricas mais tratáveis, desenvolveram uma composição da função de distribuição acumulada (quando a mesma possui inversa) com o objetivo de aumentar a assimetria e curtose em uma dada distribuição base, principalmente em distribuições simétricas.

Neste contexto, como a distribuição GEV é amplamente utilizada para modelar eventos extremos de finanças, definimos a fda da GEV transmutada a partir de:

$$F(x) = (1 + \lambda)G(x) - \lambda[G(x)]^2 \quad (3.1)$$

,

e a fdp da GEV transmutada a partir de:

$$f(x) = g(x)[(1 + \lambda) - 2\lambda G(x)] \quad (3.2)$$

.

onde $G(x)$ é a função de distribuição acumulada e $g(x)$ a função de densidade da distribuição GEV dadas em 2.2 e 2.1. Resultando em $F(x)$, a fda da distribuição transmutada com a adição do parâmetro λ , tal que $|\lambda| < 1$.

Assim sendo, as expressões da fda e a fdp da GEV transmutada, são:

Para $\gamma \neq 0$ com $x : 1 + \gamma \frac{(x-\mu)}{\sigma} > 0$, temos:

$$F(x) = e^{\left\{-\left[1+\gamma\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\gamma}}\right\}} \left[(1 + \lambda) - \lambda e^{\left\{-\left[1+\gamma\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\gamma}}\right\}} \right], \quad (3.3)$$

$$f(x) = \left[\left(1 + \gamma \frac{(x - \mu)}{\sigma} \right)^{-\frac{1}{\gamma} - 1} \frac{1}{\sigma} e^{\left\{ -[1 + \gamma \frac{(x - \mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right] \left[(1 + \lambda) - 2\lambda e^{\left\{ -[1 + \gamma \frac{(x - \mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right]. \quad (3.4)$$

Para $\gamma = 0$ com $x \in \mathbb{R}$, as expressões são:

$$F(x) = \left[e^{-e^{-\frac{(x - \mu)}{\sigma}}} \right] \left[(1 + \lambda) - \lambda e^{-e^{-\frac{(x - \mu)}{\sigma}}} \right] \quad (3.5)$$

$$f(x) = \frac{1}{\sigma} e^{-\frac{(x - \mu)}{\sigma}} e^{-e^{-\frac{(x - \mu)}{\sigma}}} \left[(1 + \lambda) - 2\lambda e^{-e^{-\frac{(x - \mu)}{\sigma}}} \right] \quad (3.6)$$

A Figura 3.1 mostra como o parâmetro λ influencia no comportamento da distribuição transmutada, em particular, para $\lambda = 0$ a transmutação retorna exatamente a distribuição base, que neste caso é a GEV. Aqui estão plotados para $\mu = 0$ e $\sigma = 1$ um gráfico com γ negativo igual à -0.5 , positivo igual à 0.5 e igual a 0 , variando os valores de λ .

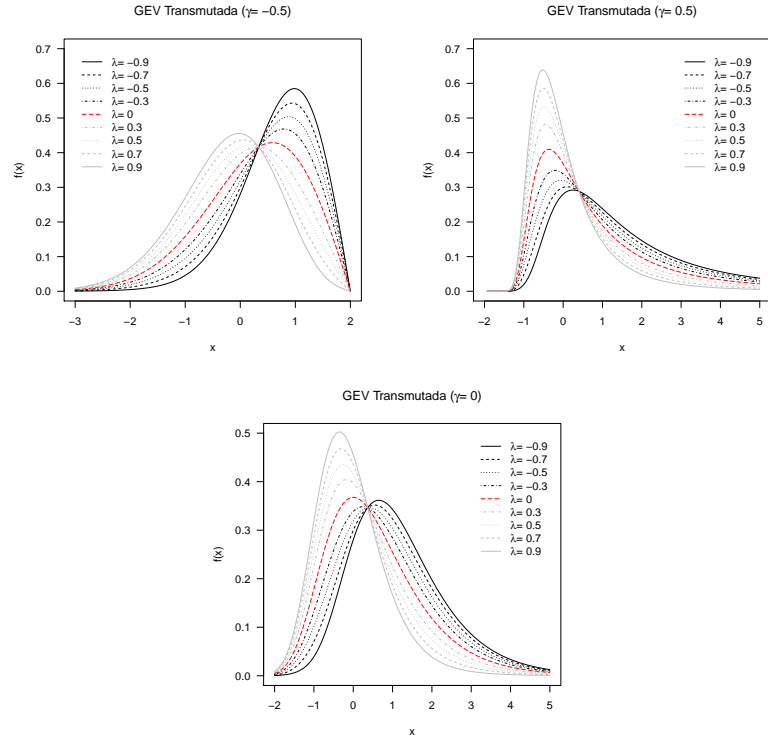


Figura 3.1 – Distribuição GEV Transmutada variando λ

Pode-se notar que o modo como o parâmetro λ modifica a distribuição ocorre de maneira diferente de acordo com os valores de γ . Observando os gráficos plotados na Figura 3.1, verifica-se que para o caso em que γ recebeu valor negativo igual a -0.5 o

parâmetro λ definiu a altura da curva de distribuição, sendo que quanto maior seu valor absoluto mais alto o ponto de máximo da função e mais pesada fica a sua calda, ocorrendo que, para os valores negativos de λ as curvas são mais altas que para os valores positivos. Da mesma forma acontece quando $\gamma = 0$, mudando apenas o fato de que as curvas mais altas estão ligadas aos valores positivos de λ . Em contra partida, quando γ é positivo, as alturas seguem exatamente o valor real de λ , atendendo a relação descrita por, quanto maior o valor de λ , maior a altura da função e maior o peso de sua cauda.

3.2 MOMENTOS

Nesta seção serão apresentados os momentos para distribuição GEV transmutada. Os momentos são medidas capazes de caracterizar uma distribuição de probabilidade. Sendo os quatro primeiros momentos indicativos de tendência central, dispersão, assimetria e curtose nesta ordem. Para a obtenção de uma fórmula geral para os momentos, calcula-se a seguir o k -ésimo momento de ordem ($E(X^k) = \int x^k f(x) dx$) da distribuição GEV transmutada para $\gamma \neq 0$ e para o caso de $\gamma = 0$ será calculada a função geratriz de momentos ($M_X(t) = E(e^{tX})$).

Para $\gamma > 0$ e atentando-se à restrição de que a expressão $1 + \gamma \frac{(x-\mu)}{\sigma}$ deve ser positiva, tem-se que o domínio da distribuição GEV transmutada se dá por $x \geq \mu - \frac{\sigma}{\gamma}$, assim:

$$E(X^k) = \int_{\mu - \frac{\sigma}{\gamma}}^{\infty} x^k \left[\left(1 + \gamma \frac{(x-\mu)}{\sigma} \right)^{-\frac{1}{\gamma}-1} \right] \left[\frac{e^{\left\{ -[1 + \gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}}}{\sigma} \right] \left[(1 + \lambda) - 2\lambda e^{\left\{ -[1 + \gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right] dx$$

Fazendo a distributiva do último termo, $\left[(1 + \lambda) - 2\lambda e^{\left\{ -[1 + \gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right]$, e separando a expressão na diferença entre duas integrais, nas quais as constantes serão postas como multiplicadoras, chega-se em:

$$\begin{aligned} E(X^k) &= \frac{(1 + \lambda)}{\sigma} \int_{\mu - \frac{\sigma}{\gamma}}^{\infty} x^k \left[\left(1 + \gamma \frac{(x-\mu)}{\sigma} \right)^{-\frac{1}{\gamma}-1} \right] \left[e^{\left\{ -[1 + \gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right] dx \\ &\quad - \frac{2\lambda}{\sigma} \int_{\mu - \frac{\sigma}{\gamma}}^{\infty} x^k \left[\left(1 + \gamma \frac{(x-\mu)}{\sigma} \right)^{-\frac{1}{\gamma}-1} \right] \left[\left(e^{\left\{ -[1 + \gamma \frac{(x-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right)^2 \right] dx \end{aligned}$$

Para resolver cada uma das integrais utiliza-se a substituição dada por $y = 1 + \gamma \frac{(x-\mu)}{\sigma}$, logo $y \geq 0$ e $dy = \frac{\gamma}{\sigma} dx$. E para substituir também o termo x^k , o x é isolado na expressão de substituição e elevado à k^a potência, sendo este polinômio simplificado usando binômio de newton de modo que $x^k = \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} y \right)^i$. Então:

$$\begin{aligned} E(X^k) &= \frac{(1+\lambda)}{\sigma} \int_0^\infty \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} y \right)^i y^{-\frac{1}{\gamma}-1} e^{-y^{-\frac{1}{\gamma}} \frac{\sigma}{\gamma}} dy \\ &\quad - \frac{2\lambda}{\sigma} \int_0^\infty \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} y \right)^i y^{-\frac{1}{\gamma}-1} e^{-2y^{-\frac{1}{\gamma}} \frac{\sigma}{\gamma}} dy \end{aligned}$$

Retirando da integral o que não se relaciona com y , obtém-se a expressão:

$$\begin{aligned} E(X^k) &= \frac{(1+\lambda)}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \int_0^\infty y^{i-\frac{1}{\gamma}-1} e^{-y^{-\frac{1}{\gamma}}} dy \\ &\quad - \frac{2\lambda}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \int_0^\infty y^{i-\frac{1}{\gamma}-1} e^{-2y^{-\frac{1}{\gamma}}} dy \end{aligned}$$

Neste momento, com o objetivo de resolver as integrais pelo método da função gama generalizada, é necessário que as expressões $i - \frac{1}{\gamma} - 1$ e $-\frac{1}{\gamma}$ sejam positivas. Para tanto, como esta é a situação em que γ é sempre maior que zero, utiliza-se uma nova substituição dada por $u = y^{-1}$ e $du = -y^{-2} dy$, com $u \in (\infty, 0]$. Com isso, o resultado é dado por:

$$\begin{aligned} E(X^k) &= \frac{(1+\lambda)}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \int_\infty^0 u^{-i+\frac{1}{\gamma}+1} e^{-u^{\frac{1}{\gamma}}} - u^{-2} du \\ &\quad - \frac{2\lambda}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \int_\infty^0 u^{-i+\frac{1}{\gamma}+1} e^{-2u^{\frac{1}{\gamma}}} - u^{-2} du \end{aligned}$$

Com essa substituição, é garantido que o expoente que anteriormente era $-\frac{1}{\gamma}$ se torne sempre positivo, dado que $\gamma > 0$. E o sinal negativo ocasionado pela substituição inverte os limites de integração.

$$\begin{aligned} E(X^k) &= \left[\frac{(1+\lambda)}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \right] \int_0^\infty u^{-i+\frac{1}{\gamma}-1} e^{-u^{\frac{1}{\gamma}}} du \\ &\quad - \left[\frac{2\lambda}{\sigma} \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^{i+1} \right] \int_0^\infty u^{-i+\frac{1}{\gamma}-1} e^{-2u^{\frac{1}{\gamma}}} du \end{aligned}$$

Agora, assumindo que o expoente $-i + \frac{1}{\gamma} - 1$ também seja positivo, ou seja, usando a restrição $i \leq \frac{1}{\gamma} - 1$ que indica que a expressão só será capaz de calcular momentos até $\frac{1}{\gamma} - 1$ (que só é positivo inteiro quando $\gamma \in (0, \frac{1}{2}]$), usa-se a função gama para resolver as integrais (função gama: $\int_0^\infty x^a e^{-bx^c} dx = \frac{\Gamma(\frac{a+1}{c})}{cb^{\frac{a+1}{c}}}$). Então, o resultado final para os momentos da distribuição GEV transmutada com $\gamma > 0$ é dado por:

$$E(X^k) = \sum_{i=0}^k C_k^i \left(\mu - \frac{\sigma}{\gamma} \right)^{k-i} \left(\frac{\sigma}{\gamma} \right)^i [\Gamma(1 - \gamma i) (1 + \lambda - 2^{\gamma i} \lambda)] \quad (3.7)$$

Para $\gamma < 0$, o que muda inicialmente são os limites de integração, que vão de $-\infty$ até $(\mu - \frac{\sigma}{\gamma})$, porém no decorrer das substituições o resultado fica exatamente igual ao encontrado para $\gamma > 0$.

Portanto, para obtenção da média da distribuição GEV transmutada com $\gamma \neq 0$ substitui-se na equação 3.7 o valor $k = 1$, lembrando que a expressão só é válida para $|\gamma| \leq 0.5$.

$$E(X) = \left(\mu - \frac{\sigma}{\gamma} \right) + \left(\frac{\sigma}{\gamma} \right) [\Gamma(1 - \gamma) (1 + \lambda - 2^\gamma \lambda)]$$

Para o cálculo da variância é preciso conhecer o segundo momento ($k = 2$) da distribuição. Dado por:

$$\begin{aligned} E(X^2) &= \left(\mu - \frac{\sigma}{\gamma} \right)^2 + 2 \left(\mu - \frac{\sigma}{\gamma} \right) \frac{\sigma}{\gamma} [\Gamma(1 - \gamma) (1 + \lambda - 2^\gamma \lambda)] \\ &+ \left(\frac{\sigma}{\gamma} \right)^2 [\Gamma(1 - 2\gamma) (1 + \lambda - 2^{2\gamma} \lambda)] \end{aligned}$$

Substituindo o resultado na fórmula da variância ($Var(X) = E(X^2) - [E(X)]^2$), e manipulando a expressão para simplificação, obtém-se:

$$Var(X) = \left(\frac{\sigma}{\gamma} \right)^2 [\Gamma(1 - 2\gamma) (1 + \lambda - 2^{2\gamma} \lambda) - \Gamma^2(1 - \gamma) (1 + \lambda - 2^\gamma \lambda)^2]$$

Agora, para o caso $\gamma = 0$, calcula-se abaixo a função geradora de momentos ($M_X(t) = E(e^{tX}) = \int e^{tx} f_X(x) dx$), com o objetivo de encontrar também, a média e variância para este caso particular da distribuição GEV transmutada.

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-\frac{(x-\mu)}{\sigma}} e^{-e^{-\frac{(x-\mu)}{\sigma}}}}{\sigma} \left[(1 + \lambda) - 2\lambda e^{-e^{-\frac{(x-\mu)}{\sigma}}} \right] dx$$

Para resolver a integral, usaremos a substituição dada por $u = e^{-\frac{(x-\mu)}{\sigma}}$, com u no intervalo $(\infty, 0)$, desta forma $x = \mu - \sigma \log(u)$ e $dx = \frac{-\sigma}{u} du$, assim reescreve-se a expressão

como:

$$E(e^{tX}) = \int_{-\infty}^0 e^{t(\mu - \sigma \log(u))} \frac{ue^{-u}}{\sigma} [1 + \lambda - 2\lambda e^{-u}] \frac{-\sigma}{u} du$$

Neste momento, é possível cancelar dentro da multiplicação u e σ restando do cancelamento o (-1) que será o fator que inverterá os limites de integração, também abriremos o expoente $t(\mu - \sigma \log(u))$ separando em duas bases.

$$E(e^{tX}) = \int_0^{\infty} e^{t\mu} e^{\log(u^{-t\sigma})} e^{-u} [(1 + \lambda) - 2\lambda e^{-u}] du$$

Retirando os termos que não dependem de u e separando as integrais após a distributiva, tem-se como resultado:

$$E(e^{tX}) = e^{t\mu} \left[(1 + \lambda) \int_0^{\infty} u^{-t\sigma} e^{-u} du - 2\lambda \int_0^{\infty} u^{-t\sigma} e^{-2u} du \right]$$

Desta forma pode-se resolver as integrais, com algumas limitações, da mesma maneira que foram resolvidas as integrais para o caso $\gamma \neq 0$, usando a função gama generalizada restringindo $t < \frac{1}{\sigma}$. O resultado é:

$$M_X(t) = E(e^{tX}) = e^{t\mu} \Gamma(1 - t\sigma) [(1 + \lambda) - 2^{t\sigma} \lambda] \quad (3.8)$$

Com a função geradora de momentos é possível gerar o momento de ordem n calculando a n -ésima derivada da função aplicada no ponto zero. A média, portanto, seria a primeira derivada $E(X) = \frac{\partial}{\partial t} M_X(t) |_{t=0}$.

$$\begin{aligned} \frac{\partial}{\partial t} M_X(t) &= \left\{ [\mu e^{t\mu} \Gamma(1 - t\sigma) + e^{t\mu} \Gamma'(1 - t\sigma)(-\sigma)] [(1 + \lambda) - 2^{t\sigma} \lambda] \right. \\ &\quad \left. + e^{t\mu} \Gamma(1 - t\sigma) [-\lambda (2^{t\sigma} \log(2)\sigma)] \right\} \Big|_{t=0} \end{aligned}$$

Muitos termos dessa equação obtida são iguais a 1 quando $t = 0$, assim a expressão final resultante para a média da distribuição GEV transmutada quando $\gamma = 0$ é:

$$E(X) = \mu + \sigma C - \lambda \sigma \log(2)$$

$$\text{onde } C = \lim_{s \rightarrow \infty} \left(\sum_{m=1}^s \frac{1}{m} - \ln(s) \right) = 0.577215$$

Novamente, para o cálculo da variância precisamos obter o segundo momento que, em se tratando de função geradora de momentos é igual à derivada de segunda ordem

no ponto $t = 0$, para que possamos substituir na fórmula da variância ($Var(X) = E(X^2) - [E(X)]^2$), então:

$$Var(X) = \sigma^2 \left[\frac{\pi^2}{6} - \lambda(1 + \lambda) \log^2(2) \right]$$

3.3 MEDIDAS DE CONFIABILIDADE

A confiabilidade para um certo tempo T , sendo T uma variável aleatória GEV transmutada é dada pela seguinte fórmula: $R(t) = 1 - F(t)$. O resultado da confiabilidade neste estudo é:

$$R(t) = 1 - \left[e^{\left\{ -\left[1 + \gamma \frac{(t-\mu)}{\sigma} \right]^{-\frac{1}{\gamma}} \right\}} \left(1 + \lambda - \lambda e^{\left\{ -\left[1 + \gamma \frac{(t-\mu)}{\sigma} \right]^{-\frac{1}{\gamma}} \right\}} \right) \right]$$

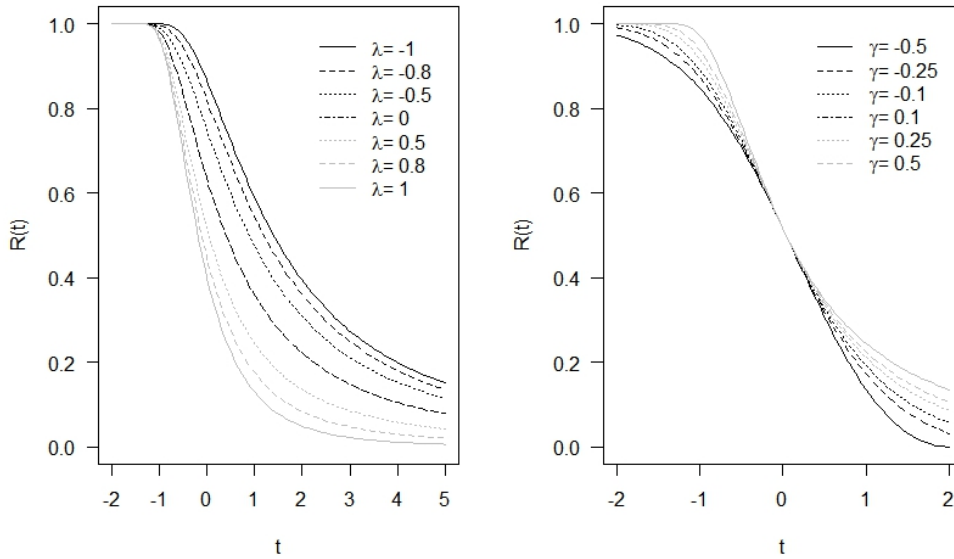


Figura 3.2 – Função de Sobrevivência da GEV Transmutada

A Figura 3.2 ilustra o comportamento da confiabilidade da distribuição GEV transmutada variando-se valores de λ e γ .

Em estatística, a função de probabilidade condicional de falha, descreve a probabilidade instantânea de ocorrência do evento "sobrevivência" até um tempo arbitrário.

Esta então seria outra característica de interesse para uma amostra aleatória da distribuição GEV transmutada. E é dada por $h(t) = \frac{f(t)}{R(t)}$:

$$h(t) = \frac{\left[\left(1 + \gamma \frac{(t-\mu)}{\sigma} \right)^{-\frac{1}{\gamma}-1} \frac{e^{\left\{ -[1+\gamma \frac{(t-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}}}{\sigma} \right] \left[(1 + \lambda) - 2\lambda e^{\left\{ -[1+\gamma \frac{(t-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right]}{1 - \left[e^{\left\{ -[1+\gamma \frac{(t-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \left(1 + \lambda - \lambda e^{\left\{ -[1+\gamma \frac{(t-\mu)}{\sigma}]^{-\frac{1}{\gamma}} \right\}} \right) \right]}$$

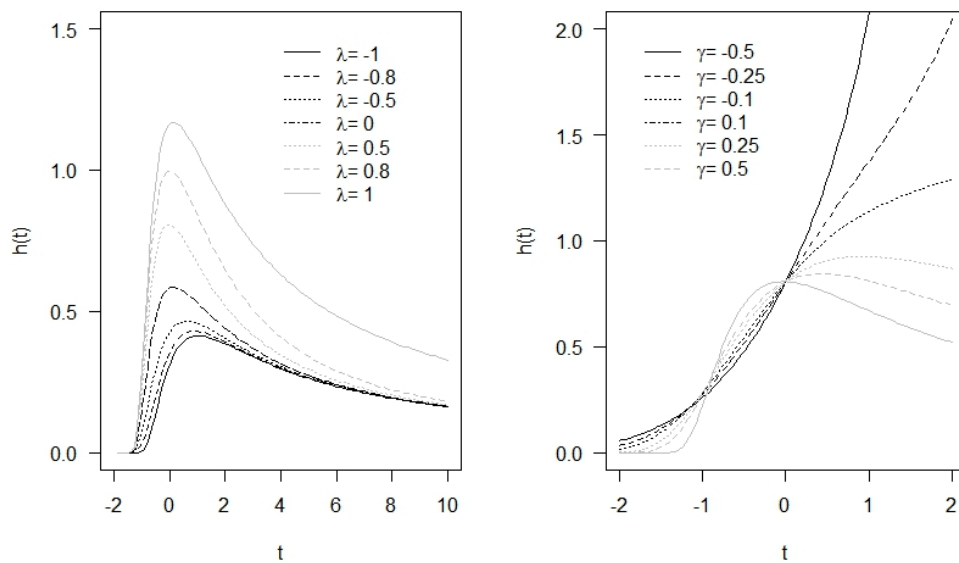


Figura 3.3 – Função de risco da GEV Transmutada

A Figura 3.3 ilustra o comportamento da função risco da distribuição GEV transmutada variando-se valores de λ e γ .

3.4 QUANTIS

A função quantil é obtida a partir da inversa da função de distribuição acumulada e retorna o valor associado à uma dada probabilidade acumulada e também pode ser usada para simular dados que seguem a distribuição.

Sendo assim, a seguir serão dados os passos para obtenção da função quantil da distribuição GEV transmutada, onde por questão de notação $F(x)$ passa a ser denotado por Q .

Para $\gamma \neq 0$

$$F(x) = Q = (1 + \lambda)e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} - \lambda \left[e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} \right]^2$$

Com objetivo de completar quadrados e facilitar os cálculos, o primeiro passo feito foi multiplicar os dois lados da igualdade por (-4λ) e logo em seguida somar em ambos os lados $(1 + \lambda)^2$, assim a equação gerada é:

$$(1 + \lambda)^2 - 4\lambda Q = (1 + \lambda)^2 - 4\lambda(1 + \lambda)e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} + \left[2\lambda e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} \right]^2$$

A partir deste ponto já é possível completar quadrados na expressão à direita da igualdade, $[(a - b)^2 = a^2 - 2ab + b^2]$, onde $a = (1 + \lambda)$ e $b = 2\lambda e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}}$.

$$(1 + \lambda)^2 - 4\lambda Q = \left[(1 + \lambda) - 2\lambda e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} \right]^2$$

Para os próximos passos deve-se aplicar raiz quadrada, em seguida, subtrair $(1 + \lambda)$ e dividir por -2λ , aplicar a função \ln e multiplicar (-1) , tudo isso em ambos os lados da igualdade:

$$\begin{aligned} \sqrt{(1 + \lambda)^2 - 4\lambda Q} &= (1 + \lambda) - 2\lambda e^{\left\{-\left[1 + \gamma \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\sigma}}\right\}} \\ -\ln \left(\frac{1 + \lambda - \sqrt{(1 + \lambda)^2 - 4\lambda Q}}{2\lambda} \right) &= \left[1 + \gamma \frac{(x - \mu)}{\sigma} \right]^{-\frac{1}{\sigma}} \end{aligned}$$

Neste momento eleva-se à $-\gamma$ e então facilmente pode-se isolar x na equação:

$$\left[-\ln \left(\frac{1 + \lambda - \sqrt{(1 + \lambda)^2 - 4\lambda Q}}{2\lambda} \right) \right]^{-\gamma} = 1 + \gamma \frac{(x - \mu)}{\sigma}$$

O resultado final para a função quantil da distribuição GEV transmutada é dado por:

$$\frac{\sigma \left\{ -1 + \left[-\ln \left(\frac{1 + \lambda - \sqrt{(1 + \lambda)^2 - 4\lambda Q}}{2\lambda} \right) \right]^{-\gamma} \right\}}{\gamma} + \mu = x \quad (3.9)$$

Para o caso particular em que $\gamma = 0$, os passos são basicamente os mesmos. No início multiplica-se a expressão por -4λ , soma $(1 + \lambda)^2$, completa-se quadrados, aplica-se

a raiz quadrada, subtrai $(1 + \lambda)$ e continua-se realizando passos até que x esteja isolado na igualdade. Obtendo:

$$Q = (1 + \lambda)e^{-e^{-\frac{(x-\mu)}{\sigma}}} - \lambda \left[e^{-e^{-\frac{(x-\mu)}{\sigma}}} \right]^2$$

$$\sigma \left\{ -\ln \left[-\ln \left(\frac{1 + \lambda - \sqrt{(1 + \lambda)^2 - 4\lambda Q}}{2\lambda} \right) \right] \right\} + \mu = x \quad (3.10)$$

As expressões quantis são muito utilizadas em simulações de dados. Neste trabalho, no Capítulo 4, as simulações de dados serão feitas via método da transformação inversa, sendo utilizadas as funções quantis.

3.5 ESTATÍSTICAS DE ORDEM

Sejam $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$ as estatísticas de ordem de uma amostra aleatória $X_1, X_2, X_3, \dots, X_n$ obtida de uma população que segue uma distribuição $f(x)$, então a função de distribuição de probabilidade da j -ésima estatística de ordem é dada por:

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}$$

para $j = 1, 2, 3, \dots, n$

Aplicando as funções de densidade de probabilidade 2.1 e de distribuição acumulada 2.2 da distribuição GEV obtém-se $g_{X_{(j)}}$, a fdp da j -ésima estatística de ordem da GEV.

Para $\gamma \neq 0$, foi utilizada a substituição $w = 1 + \gamma \frac{(x-\mu)}{\sigma}$ com o objetivo de diminuir a expressão, então:

$$g_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \left[e^{-w^{-\frac{1}{\gamma}}} \frac{1}{\sigma} w^{-\frac{1}{\gamma}-1} \right] \left[e^{-w^{-\frac{1}{\gamma}}} \right]^{j-1} \left[1 - e^{-w^{-\frac{1}{\gamma}}} \right]^{n-j}$$

Para simplificar, o expoente $(j-1)$ do termo $\left[e^{-w^{-\frac{1}{\gamma}}} \right]$ é colocado como multiplicador junto o primeiro expoente:

$$g_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \left[\frac{1}{\sigma} w^{-\frac{1}{\gamma}-1} e^{-w^{-\frac{1}{\gamma}}} \right] \left[e^{-(j-1)w^{-\frac{1}{\gamma}}} \right] \left[1 - e^{-w^{-\frac{1}{\gamma}}} \right]^{n-j}$$

Desta forma é possível juntar as exponenciais obtendo a fórmula geral para a j -ésima estatística de ordem da GEV:

$$g_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \frac{1}{\sigma} w^{-\frac{1}{\gamma}-1} e^{-jw^{-\frac{1}{\gamma}}} \left[1 - e^{-w^{-\frac{1}{\gamma}}} \right]^{n-j} \quad (3.11)$$

Como a distribuição GEV está interessada em modelar eventos máximos ou mínimos, a seguir estão calculadas a primeira e n-ésima estatísticas de ordem. Para a estatística de ordem n , substitui-se na fórmula geral 3.5 $j = n$. Portanto, a n-ésima estatística de ordem, que representa a distribuição do máximo, da distribuição GEV é dada por:

$$g_{X_{(n)}}(x) = \frac{n}{\sigma} w^{-\frac{1}{\gamma}-1} e^{-nw^{-\frac{1}{\gamma}}}$$

Para a estatística de ordem 1, substitui-se na fórmula geral $j=1$. O resultado gera a distribuição do mínimo de uma distribuição GEV:

$$g_{X_{(1)}}(x) = \frac{n}{\sigma} w^{-\frac{1}{\gamma}-1} e^{-w^{-\frac{1}{\gamma}}} \left[1 - e^{-w^{-\frac{1}{\gamma}}} \right]^{n-1}$$

Quando $\gamma = 0$ a fórmula geral para a estatística de ordem é dada a seguir, em que foi utilizada a substituição $v = e^{-\frac{(x-\mu)}{\sigma}}$:

$$g_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \frac{ve^{-vj}}{\sigma} (1 - e^{-v})^n - j \quad (3.12)$$

As funções das estatísticas de n-ésima e primeira ordem, correspondentes ao máximo e mínimo da distribuição GEV para $\gamma = 0$ estão descritas a seguir:

$$g_{X_{(n)}}(x) = n \frac{ve^{-nv}}{\sigma} \quad ; \quad g_{X_{(1)}}(x) = n \frac{ve^{-v}}{\sigma} (1 - e^{-v})^{n-1}.$$

Em particular, observa-se que a estatística de ordem 1, quando $n = 2$, é exatamente a f.d.p. da GEV transmutada para $\lambda = 1$. Esta característica ocorre de maneira semelhante para a segunda estatística de ordem da GEV, onde, para $n = 2$, tem-se que é exatamente igual a f.d.p. da GEV transmutada para $\lambda = -1$. Como nestes casos $n=2$, a estatística de ordem 1 representa a distribuição do mínimo e a estatística de ordem 2 o máximo.

Agora, será calculada a j-ésima estatística de ordem para uma amostra aleatória da distribuição GEV transmutada no caso em que $\gamma \neq 0$, permanecendo a substituição utilizada anteriormente dada por $w = 1 + \gamma \frac{(x-\mu)}{\sigma}$. O resultado obtido é:

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(j-1)!(n-j)!} \left[\left(w^{-\frac{1}{\gamma}-1} \frac{e^{-w^{-\frac{1}{\gamma}}}}{\sigma} \right) \left(1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right] \\
&\times \left[e^{-w^{-\frac{1}{\gamma}}} \left(1 + \lambda - \lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right]^{j-1} \left[1 - e^{-w^{-\frac{1}{\gamma}}} \left(1 + \lambda - \lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right]^{n-j}
\end{aligned} \tag{3.13}$$

Novamente, para encontrar as estatísticas de ordem n e 1 substitui-se o valor de j na equação 3.13, encontrando as distribuições do máximo e mínimo da GEV transmutada:

$$\begin{aligned}
f_{X_{(n)}}(x) &= n \left[\left(w^{-\frac{1}{\gamma}-1} \frac{e^{-w^{-\frac{1}{\gamma}}}}{\sigma} \right) \left(1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right] \left[e^{-w^{-\frac{1}{\gamma}}} \left(1 + \lambda - \lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right]^{n-1} \\
f_{X_{(1)}}(x) &= n \left[\left(w^{-\frac{1}{\gamma}-1} \frac{e^{-w^{-\frac{1}{\gamma}}}}{\sigma} \right) \left(1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right] \left[1 - e^{-w^{-\frac{1}{\gamma}}} \left(1 + \lambda - \lambda e^{-w^{-\frac{1}{\gamma}}} \right) \right]^{n-1}
\end{aligned}$$

Para o caso de $\gamma = 0$, repetindo procedimentos semelhantes aos realizados anteriormente, o resultado obtido para a j -ésima estatística de ordem é dado pela expressão a seguir, em que $v = e^{-\frac{(x-\mu)}{\sigma}}$:

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(j-1)!(n-j)!} \frac{ve^{-v}}{\sigma} [(1+\lambda) - 2\lambda e^{-v}] \\
&\times [(1+\lambda - \lambda e^{-v})^{j-1}] [1 - (e^{-v}(1+\lambda - \lambda e^{-v}))]^{n-j}
\end{aligned} \tag{3.14}$$

Obtemos então as estatísticas n -ésima e de primeira ordem, substituindo j pelos valores n e 1, respectivamente, na fórmula 3.14 resultando na distribuição de máximo e mínimo da GEV transmutada para $\gamma = 0$:

$$f_{X_{(n)}}(x) = n \frac{ve^{-nv}}{\sigma} [(1+\lambda) - 2\lambda e^{-v}] [(1+\lambda - \lambda e^{-v})^{n-1}]$$

$$f_{X_{(1)}}(x) = n \frac{ve^{-v}}{\sigma} [(1+\lambda) - 2\lambda e^{-v}]$$

Capítulo 4

Estimação e Simulação

Neste capítulo serão apresentadas simulações da distribuição GEV transmutada realizadas computacionalmente e seus resultados. Para tanto foi utilizado o software estatístico RStudio em sua versão 3.2.4. O objetivo foi gerar amostras que seguem a distribuição em estudo com parâmetros conhecidos para então aplicar o método de estimação de máxima verossimilhança e avaliar a eficiência do mesmo.

4.1 ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA (MV)

O método utilizado para a estimação dos parâmetros procura, baseado nos resultados obtidos por amostra, o parâmetro (ou o conjunto de parâmetros) que maximizam a função de verossimilhança (L) de uma distribuição de probabilidades. Tal função é definida da seguinte forma, se x_1, x_2, \dots, x_n for uma amostra aleatória (independente e identicamente distribuída), então:

$$L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta)$$

Usualmente, utiliza-se a função conhecida como log-verossimilhança, que nada mais é que a aplicação da função logaritmo na função de máxima verossimilhança.

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_n) = \log(L(\theta; x_1, x_2, \dots, x_n)) = \sum_{i=1}^n \log(f(x_i|\theta))$$

Para a distribuição GEV transmutada as funções de máxima log-verossimilhança estão descritas abaixo para o caso de $\gamma = 0$ e $\gamma \neq 0$ respectivamente:

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_n) = - \sum_{i=1}^n \left[\log(\sigma) + \frac{x_i - \mu}{\sigma} + e^{-\frac{(x_i - \mu)}{\sigma}} \right] + \sum_{i=1}^n \log(1 + \lambda - 2\lambda e^{-e^{-\frac{(x_i - \mu)}{\sigma}}})$$

$$\begin{aligned}\mathcal{L}(\theta; x_1, x_2, \dots, x_n) &= -n \log(\sigma) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^n \log \left(1 + \frac{\gamma}{\sigma}(x_i - \mu)\right) \\ &\quad - \sum_{i=1}^n \left(1 + \frac{\gamma}{\sigma}(x_i - \mu)\right)^{-\frac{1}{\gamma}} + \sum_{i=1}^n \log \left(1 + \lambda - 2\lambda e^{\left\{-(1 + \frac{\gamma}{\sigma}(x_i - \mu))^{-\frac{1}{\gamma}}\right\}}\right)\end{aligned}$$

Os estimadores de máxima verossimilhança são obtidos pela solução do sistema de equações dados pelas derivadas da função log-verossimilhança em relação a cada um dos parâmetros a serem estimados. A seguir estão descritos os sistemas gerados pela distribuição GEV transmutada para a obtenção das estimativas.

Para $\gamma = 0$ foi utilizada, no sistema de equações, a substituição dada por $w = \frac{x_i - \mu}{\sigma}$ com o objetivo apenas de diminuir a expressão. Temos então:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mu} &= n - \sum_{i=1}^n e^{-w} + 2\lambda \sum_{i=1}^n \frac{e^{-w} e^{-e^{-w}}}{1 + \lambda - 2\lambda e^{-e^{-w}}} = 0 \\ \frac{\partial \mathcal{L}}{\partial \sigma} &= -n + \sum_{i=1}^n (x_i - \mu) [1 - e^{-w}] + \frac{2\lambda}{\sigma} \sum_{i=1}^n \frac{(x_i - \mu) e^{-w} e^{-e^{-w}}}{1 + \lambda - 2\lambda e^{-e^{-w}}} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \sum_{i=1}^n \frac{1 - 2e^{-e^{-w}}}{1 + \lambda - 2\lambda e^{-e^{-w}}} = 0\end{aligned}$$

Também foi feita uma substituição para quando $\gamma \neq 0$ com o mesmo objetivo e esta se dá da seguinte forma $w = 1 + \frac{\gamma}{\sigma}(x_i - \mu)$, assim:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mu} &= \sum_{i=1}^n \frac{1 + \gamma}{\sigma + \gamma(x_i - \mu)} - \sum_{i=1}^n \frac{w^{-\frac{1}{\gamma}-1}}{\sigma} + \sum_{i=1}^n \frac{w^{-\frac{1}{\gamma}-1} 2\lambda e^{-w^{-\frac{1}{\gamma}}}}{\sigma \left(1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}}\right)} = 0 \\ \frac{\partial \mathcal{L}}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(1 + \gamma)(x_i - \mu)}{\sigma[\sigma + \gamma(x_i - \mu)]} - \sum_{i=1}^n \frac{(x_i - \mu) w^{-\frac{1}{\gamma}-1}}{\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu) w^{-\frac{1}{\gamma}-1} 2\lambda e^{-w^{-\frac{1}{\gamma}}}}{\sigma^2 [1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}}] } = 0 \\ \frac{\partial \mathcal{L}}{\partial \gamma} &= \sum_{i=1}^n \left[\frac{\log(w)}{\gamma^2} + \frac{\left(\frac{1}{\gamma}\right)(x_i - \mu)}{\sigma w} \right] - \sum_{i=1}^n w^{-\frac{1}{\gamma}} \left[\frac{\log(w)}{\gamma^2} - \frac{(x_i - \mu)}{\gamma \sigma w} \right] \\ &\quad + \sum_{i=1}^n w^{-\frac{1}{\gamma}} \left[\frac{\log(w)}{\gamma^2} - \frac{(x_i - \mu)}{\gamma \sigma w} \right] \left[\frac{2\lambda e^{-w^{-\frac{1}{\gamma}}}}{1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}}} \right] = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \sum_{i=1}^n \frac{1 - 2e^{-w^{-\frac{1}{\gamma}}}}{1 + \lambda - 2\lambda e^{-w^{-\frac{1}{\gamma}}}} = 0\end{aligned}$$

Para a resolução dos sistemas de equações e portanto, para encontrar as estimativas dos parâmetros foi utilizado um método de minimização de funções, desenvolvido por Nelder, J. A. e Mead, R. (1965), através do software estatístico de programação, RStudio.

4.2 ILUSTRAÇÕES NUMÉRICAS

Para realizar as simulações de amostras foi utilizado o método da transformada inversa. Este método utiliza a função quantil (aqui apresentadas nas equações 3.9 e 3.10) para gerar valores de determinada variável aleatória X que possua função de distribuição acumulada $F_X(x)$. Os valores são gerados por $X = F_X^{-1}(U)$, em que U segue distribuição Uniforme variando de 0 à 1.

Foram geradas, em cada uma das combinações de parâmetros definidas, $M = 100$ amostras de tamanho $n = 1000$. Os parâmetros μ e σ se mantiveram constantes iguais a 0 e 1, em todas as combinações utilizadas, o parâmetro γ variou entre os valores -0.9 , -0.5 , 0 , 0.5 , e 0.9 , e por fim, λ recebeu valores igual à -0.9 , 0.1 , e 0.9 . Com isso foi possível a criação de três cenários distintos, caracterizados pelo valor de γ , sendo assim, no cenário 1 estão as combinações em que $\gamma > 0$, nos cenários 2 e 3 encontram-se $\gamma = 0$ e $\gamma < 0$ respectivamente.

Em cada um dos cenários foram calculadas as estimativas dos parâmetros (utilizando a função `optim` do R - R Core Team, 2015) para as $M = 100$ simulações geradas e o resultado para a estimação apresentado na tabela 4.1 é definido por $\bar{\theta} = \sum_{i=1}^M \frac{\hat{\theta}_i}{M}$. Os vieses $((\theta - \bar{\theta}))$ gerados pelas estimativas e os erros quadráticos médios para cada um dos parâmetros $\left(EQM = \sum_{i=1}^M \frac{(\hat{\theta}_i - \bar{\theta})^2}{M}\right)$ estão na Tabela 4.2 abaixo.

Tabela 4.1 – Parâmetros e estimativas no cenário 1 ($\gamma > 0$)

$\theta = (\mu; \sigma; \gamma; \lambda)$	$\bar{\theta} = (\bar{\mu}; \bar{\sigma}; \bar{\gamma}; \bar{\lambda})$
$\theta_1 = (0; 1; 0.9; -0.9)$	$\bar{\theta}_1 = (0.1802; 1.1515; 0.8918; -0.7034)$
$\theta_2 = (0; 1; 0.5; -0.9)$	$\bar{\theta}_2 = (0.1383; 1.0659; 0.5054; -0.7314)$
$\theta_3 = (0; 1; 0.9; 0.1)$	$\bar{\theta}_3 = (-0.1746; 0.8377; 0.9138; -0.2703)$
$\theta_4 = (0; 1; 0.5; 0.1)$	$\bar{\theta}_4 = (-0.2787; 0.8102; 0.5122; -0.3857)$
$\theta_5 = (0; 1; 0.9; 0.9)$	$\bar{\theta}_5 = (-0.3484; 0.5384; 0.6354; -0.0313)$
$\theta_6 = (0; 1; 0.5; 0.9)$	$\bar{\theta}_6 = (-0.4360; 0.5926; 0.3278; -0.1449)$

O resultado que mais chama a atenção nas Tabela 4.1 se deve ao fato de que todas as estimativas para o parâmetro λ foram negativas, e em especial, a combinação de θ_5 obteve uma estimativa para este parâmetro bastante próxima de zero, que supostamente retorna a GEV transmutada para sua forma base que é a GEV.

A Figura 4.1 abaixo ilustra as curvas regidas pelos parâmetros estimados em cor azul juntamente com as curvas geradas pelos valores reais dos parâmetros em cor vermelha. Em todas as combinações de parâmetros utilizadas o ajuste observado nos gráficos é aparentemente muito bom. Porém ao analisar a Tabela 4.2, que contém os vieses, podemos notar que algumas diferenças entre o verdadeiro valor do parâmetro e a estimativa obtida são razoavelmente significantes, apesar de resultarem em curvas extremamente semelhantes. Este fato, muito provavelmente se deve à existência de relação entre os parâmetros.

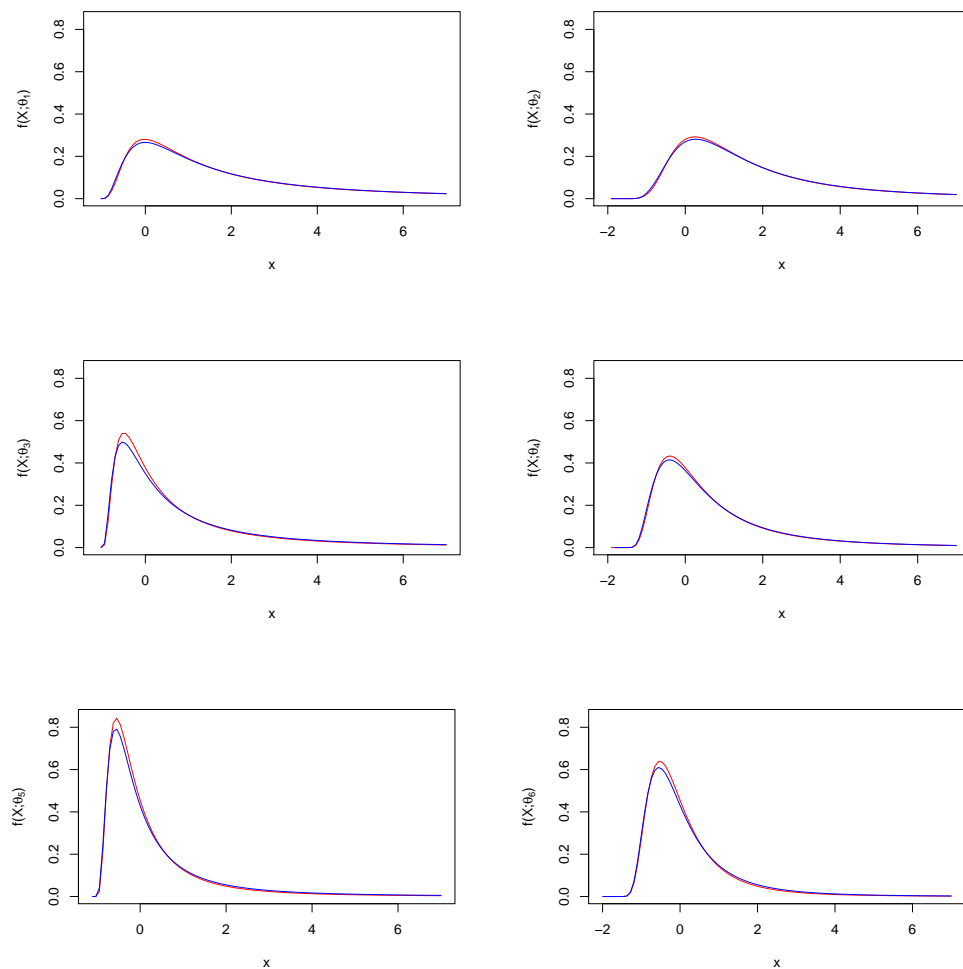


Figura 4.1 – Distribuição GEV teórica e sua estimativa para as combinações no cenário 1.

Analisando os vieses obtidos parâmetro a parâmetro, contidos na Tabela 4.2, pode-se notar que o γ foi muito bem estimado nos casos em que os valores de λ foram -0.9 e 0.1 e as piores estimativas foram encontradas para λ quando ele assumia o valor real de 0.9. A tabela apresenta também os valores do EQM para as estimativas, os baixos valores para esta medida mostram que o algoritmo obteve estimativas semelhantes em todas as $M = 100$ amostras geradas.

Tabela 4.2 – Viés e EQM no cenário 1 ($\gamma > 0$)

θ	Viés	EQM			
		$\bar{\mu}$	$\bar{\sigma}$	$\bar{\gamma}$	$\bar{\lambda}$
θ_1	(-0.1802; -0.1515; 0.0082; -0.1966)	0.0931	0.1094	0.0019	0.0891
θ_2	(-0.1383; -0.0659; -0.0054; -0.1686)	0.0615	0.0267	0.0011	0.0884
θ_3	(0.1746; 0.1623; -0.0138; 0.3703)	0.0735	0.1077	0.0052	0.2348
θ_4	(0.2787; 0.1898; -0.0122; 0.4857)	0.0428	0.0191	0.0012	0.1325
θ_5	(0.3484; 0.4616; 0.2646; 0.9313)	0.0283	0.0389	0.0122	0.2374
θ_6	(0.4360; 0.4074; 0.1722; 1.0449)	0.0375	0.0205	0.0026	0.2407

Partindo para o cenário 2, onde tem-se que $\gamma = 0$, foram realizados os mesmos procedimentos e os resultados obtidos estão descritos a seguir:

Tabela 4.3 – Parâmetros e estimativas no cenário 2 ($\gamma = 0$)

$\theta = (\mu; \sigma; \lambda)$	$\bar{\theta} = (\bar{\mu}; \bar{\sigma}; \bar{\lambda})$
$\theta_1 = (0; 1; -0.9)$	$\bar{\theta}_1 = (0.6664; 1.1299; 0.0545)$
$\theta_2 = (0; 1; 0.1)$	$\bar{\theta}_2 = (0.0502; 1.0370; 0.1775)$
$\theta_3 = (0; 1; 0.9)$	$\bar{\theta}_3 = (-0.2773; 0.8864; 0.3295)$

Tabela 4.4 – Viés e EQM no cenário 2 ($\gamma = 0$)

θ	Viés	EQM		
		$\bar{\mu}$	$\bar{\sigma}$	$\bar{\lambda}$
θ_1	(-0.6664; -0.1299; -0.9545)	0.1971	0.0258	0.4194
θ_2	(-0.0502; -0.0370; -0.0775)	0.0392	0.0087	0.0988
θ_3	(0.2773; 0.1136; 0.5705)	0.1279	0.0229	0.5355

No cenário 2, encontramos uma combinação de parâmetros, aqui definida como θ_2 , que foi muito bem estimada pelo algoritmo, gerando vieses mínimos para cada um dos parâmetros, como descrito na Tabela 4.4. No respectivo gráfico apresentado na Figura 4.2 para θ_2 , existe dificuldade na distinção entre a função real e a estimada. Já para as combinações θ_1 e θ_3 os resultados mostram estimativas não tão boas para todos os parâmetros, em especial para λ , que obteve os maiores vieses nas duas combinações. O reflexo disto pode também ser visto na Figura 4.2, em que os gráficos das respectivas combinações mostram um desajuste entre a curva com parâmetros reais em vermelho e sua estimativa em azul. Novamente, uma das estimativas para λ resultou em um valor próximo de zero, neste caso para a combinação definida por θ_1 .

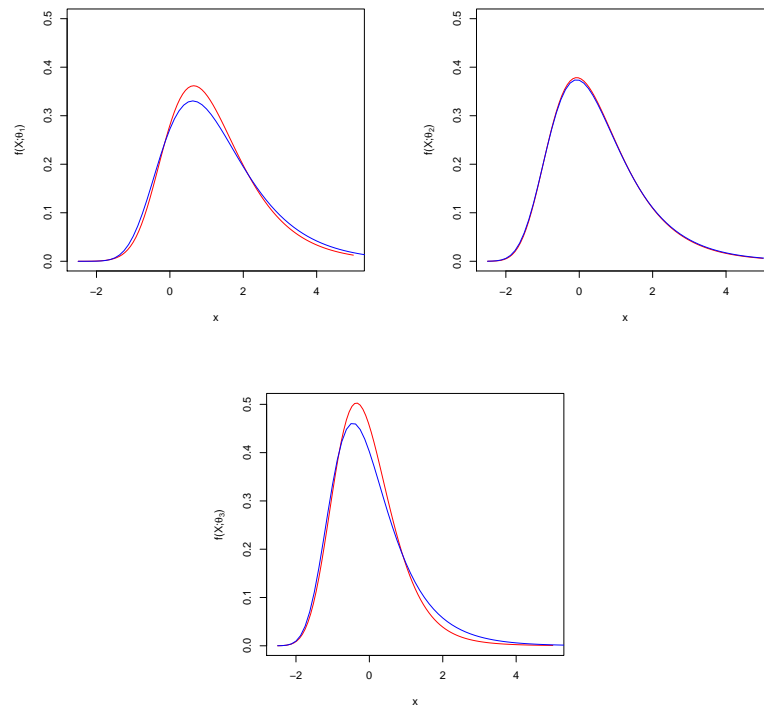


Figura 4.2 – Distribuição GEV teórica e sua estimativa para as 3 combinações no cenário 2.

Por fim, os resultados das simulações geradas para o cenário 3, caracterizado por $\gamma < 0$, são apresentados. Neste cenário, existiram alguns problemas no algoritmo de estimação para duas das combinações de parâmetros, o que levou à retirada uma das combinações e a substituição forçada de um dos parâmetros na outra. Esses casos ocorreram nas estimações que seriam $(\mu = 0, \sigma = 1, \gamma = -0.9, \lambda = -0.9)$ e $(\mu = 0, \sigma = 1, \gamma = 0.9, \lambda = 0.1)$ sendo

que a primeira configuração de parâmetros foi removida e na segunda o valor de λ foi substituído por 0.5. Dadas as novas combinações de parâmetros a serem simulados e estimados, os resultados são:

Tabela 4.5 – Parâmetros e estimativas no cenário 3 ($\gamma < 0$)

$\theta = (\mu; \sigma; \gamma; \lambda)$	$\bar{\theta} = (\bar{\mu}; \bar{\sigma}; \bar{\gamma}; \bar{\lambda})$
$\theta_1 = (0; 1; -0.5; -0.9)$	$\bar{\theta}_1 = (0.0845; 0.9606; -0.5039; -0.7625)$
$\theta_2 = (0; 1; -0.5; 0.1)$	$\bar{\theta}_2 = (-0.4996; 1.1949; -0.4797; -0.5651)$
$\theta_3 = (0; 1; -0.9; 0.5)$	$\bar{\theta}_3 = (-0.7780; 1.3864; -0.7326; -0.5554)$
$\theta_4 = (0; 1; -0.5; 0.9)$	$\bar{\theta}_4 = (-0.9229; 1.0443; -0.3653; -0.6248)$
$\theta_5 = (0; 1; -0.9; 0.9)$	$\bar{\theta}_5 = (-1.0467; 1.3478; -0.6207; -0.6079)$

Novamente todas as estimativas do parâmetro λ foram negativas, assim como ocorreu no cenário 1. A Tabela 4.6 apresenta os vieses e os erros quadráticos médios obtidos para as cinco combinações de parâmetros simuladas.

Tabela 4.6 – Viés e EQM no cenário 3 ($\gamma < 0$)

θ	Viés	EQM			
		$\bar{\mu}$	$\bar{\sigma}$	$\bar{\gamma}$	$\bar{\lambda}$
θ_1	$(-0.0845; 0.0394; 0.0039; -0.1375)$	0.0320	0.0101	0.0009	0.0792
θ_2	$(0.4996; -0.1949; -0.0203; 0.6651)$	0.1336	0.0243	0.0014	0.2299
θ_3	$(0.7780; -0.3864; -0.1674; 1.0554)$	0.0629	0.0366	0.0005	0.0653
θ_4	$(0.9229; -0.0443; -0.1347; 1.5248)$	0.0675	0.0067	0.0016	0.1520
θ_5	$(1.046; -0.3478; -0.2793; 1.5079)$	0.0380	0.0096	0.0009	0.0534

As ilustrações apresentadas na Figura 4.3 mostram que, assim como no cenário 1, apesar de o método obter estimativas por vezes distantes dos reais valores dos parâmetros, as curvas estimadas são visualmente muito satisfatórias. Mais uma vez a explicação deste fato pode ser devido a uma possível relação entre os parâmetros que infelizmente não será tratada neste trabalho.

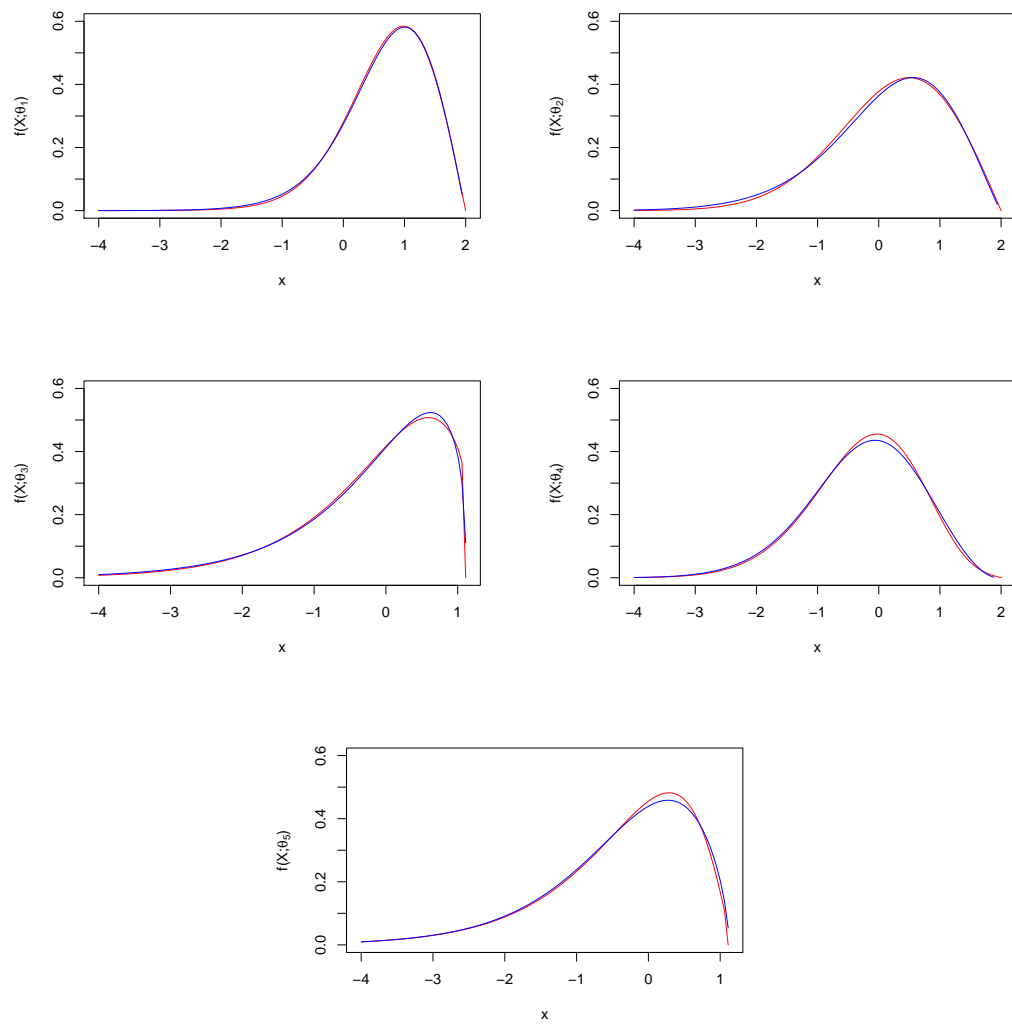


Figura 4.3 – Distribuição GEV teórica e sua estimativa para as combinações no cenário 3.

Capítulo 5

Análise em Dados Reais

Com o objetivo de aplicar os resultados obtidos até então, foram utilizados três bancos de dados reais sobre índices financeiros. O primeiro deles se trata do Ibovespa, um grande índice de ações da bolsa de valores de São Paulo que reflete o desempenho das cotações das ações mais representativas do mercado acionário brasileiro. Outro banco se refere ao S&P 500, um índice baseado em capitalizações de 500 grandes empresas do mercado de ações americano. Também foi utilizado o índice Dow Jones que é baseado na cotação das ações de 30 das maiores e mais importantes empresas dos Estados Unidos. Todos os dados utilizados podem ser obtidos através do site <http://br.investing.com/indices>.

Para cada um dos índices foram utilizados dados diários dos últimos 10 anos, a partir de 03/01/2006 até 31/10/2016. Para o Ibovespa foram obtidas 2.707 observações, para o S&P 500 foram 2768 e para o Dow Jones 2795. A fonte de dados fornece os últimos valores dos índices no dia, bem como seus valores de abertura, máximo e mínimo, sendo que, para esta análise foram utilizados os valores de abertura em todos os índices.

Por se tratar de dados financeiros, é comum o interesse em analisar o comportamento da variação dos índices. Para isso, é utilizado o retorno, que pode ser definido de algumas maneiras distintas, e neste trabalho foi utilizado o retorno composto continuamente, ou log-retorno, que é definido segundo a equação:

$$r_t = \log(P_t) - \log(P_{t-1}),$$

em que P_t corresponde ao valor no instante t .

Calculados os retornos, é possível notar que, assim como é de se esperar, no comportamento de retornos há o excesso de curtose, claramente visualizado na Figura 5.1:

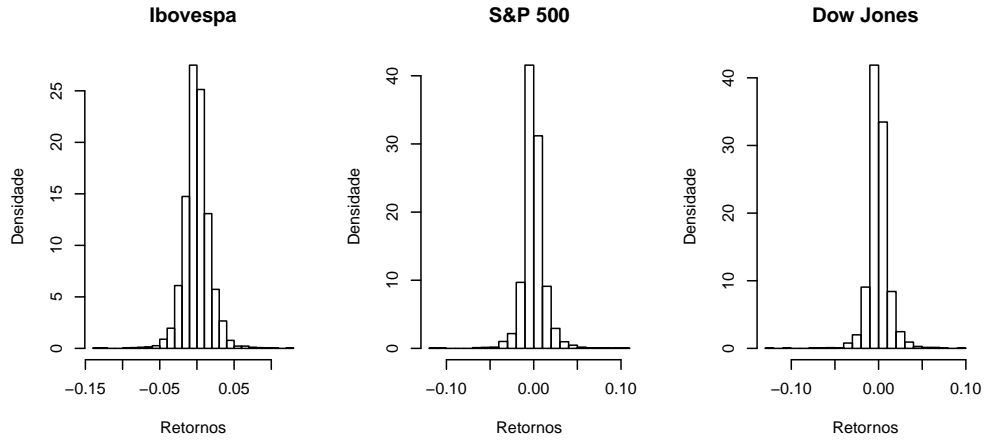


Figura 5.1 – Retornos dos índices.

Visto que a distribuição desenvolvida neste estudo modela valores extremais, foram pegos, dentro de blocos compostos por 7 dias, os valores máximos dos retornos. Esta técnica é conhecida por Block Max. Então, os novos histogramas para os dados se comportam como ilustrado na Figura 5.2:

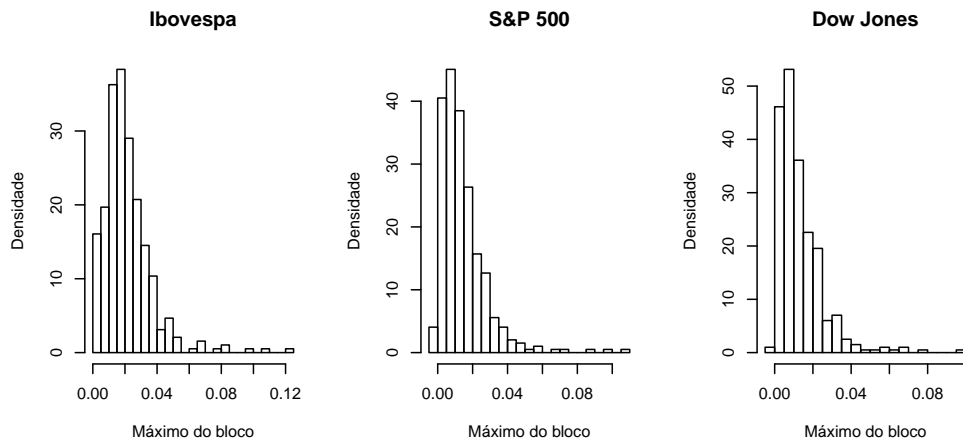


Figura 5.2 – Medidas de máximo em blocos

A seguir estão apresentadas as principais estatísticas descritivas para os três bancos de dados em questão. Na Tabela 5.1 tem-se a média ($E(X)$), variância ($Var(X)$), coeficiente de assimetria ($ca(X)$) definido por $ca(x) = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$, máximo ($\max(X)$), mínimo ($\min(X)$) e mediana ($mediana(X)$).

Tabela 5.1 – Estatísticas descritivas dos dados.

ÍNDICE	$E(X)$	$\text{Var}(X)$	$\text{ca}(X)$	$\text{max}(X)$	$\text{min}(X)$	$\text{mediana}(X)$
Ibovespa	0.0214	0.0002	2.3572	0.1208	3×10^{-6}	0.0186
S&P 500	0.0144	0.0001	2.7219	0.1055	-7×10^{-4}	0.0112
Dow Jones	0.0130	0.0001	2.5203	0.0966	-0.0011	0.0099

A partir dos bancos de dados gerados pelos máximos de blocos, é a hora de ajustar a distribuição GEV transmutada, estimando os parâmetros pelo método de máxima verossimilhança. Em seguida, para checar o ajuste analisar os qqplots para cada um deles. Abaixo, tem-se a tabela com as estimações obtidas, o histograma dos dados com a curva estimada e o qqplot para cada banco:

Tabela 5.2 – Parâmetros estimados.

ÍNDICE	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\lambda}$
IBOVESPA	0.02098	0.01500	0.28645	0.95497
S&P 500	0.01222	0.01080	0.43309	0.80260
DOW JONES	0.00782	0.00651	0.30470	0.14741

Dos resultados da estimativas, podemos observar que o parâmetro λ foi estimado em valores diferentes de 0 (zero), indicando que a distribuição de probabilidade GEV não seria mais adequada que a GEV transmutada para estes bancos de dados.

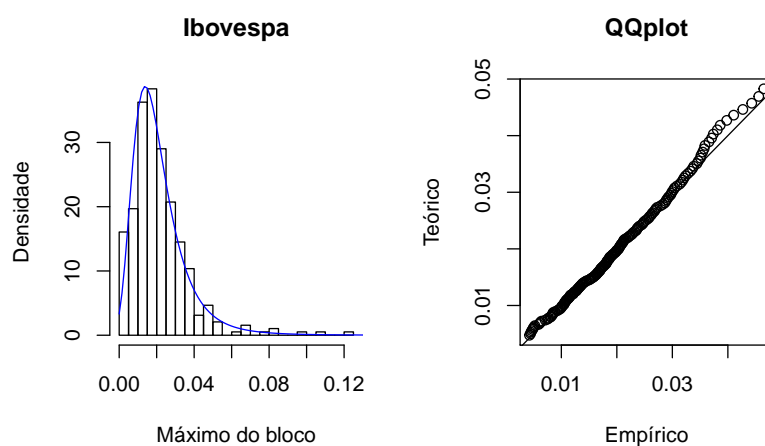


Figura 5.3 – Curva estimada para os dados do Ibovespa.

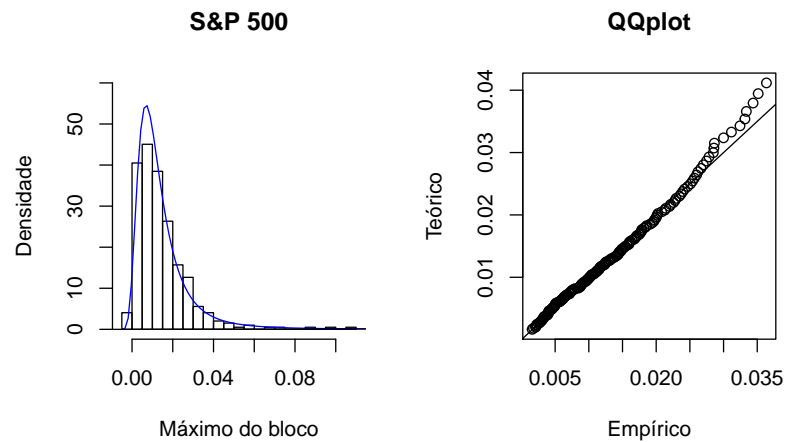


Figura 5.4 – Curva estimada para os dados do S&P 500.

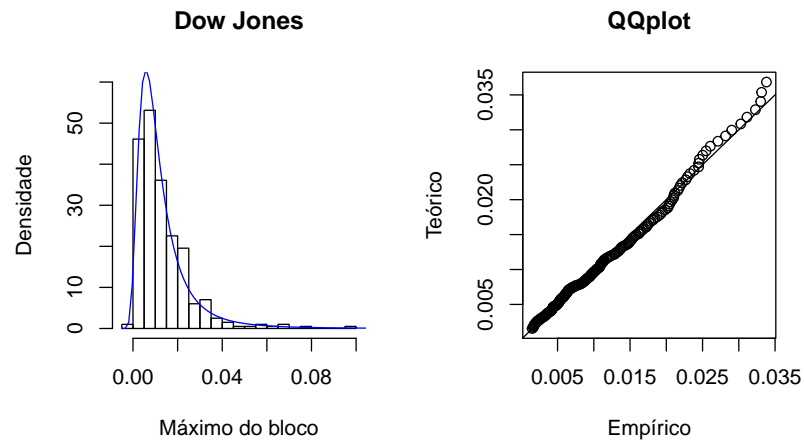


Figura 5.5 – Curva estimada para os dados do Dow Jones.

Em todos os casos podemos observar, analisando os gráficos quantil a quantil, que a distribuição GEV transmutada se ajusta muito bem como estimativa de distribuição para os dados utilizados. Para verificar o ajustamento da distribuição estimada aos dados, foi realizado o teste de Kolmogorov-Smirnov, este é um teste não paramétrico baseado na diferença absoluta máxima entre a função de distribuição acumulada estimada e a função de distribuição empírica dos dados, e também o teste de Anderson Darling. Ambos testam as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Os dados seguem a distribuição em questão;} \\ H_A : \text{Os dados não seguem a distribuição.} \end{cases}$$

Para tanto, a estatística utilizada no teste de Kolmogorov-Smirnov é dada por: $D_n = \sup_x |F_n(x) - F(x)|$. Já a estatística para o teste de Anderson Darling é definida segundo a seguinte expressão: $A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]}{F(x)[1 - F(x)]} dF(x)$. Em que, $F(x)$ representa a função de distribuição acumulada assumida para os dados e $F_n(x)$ representa a função de distribuição acumulada empírica dos dados, correspondente à proporção de valores menores ou iguais a x .

Desta forma, a Tabela 5.3 estão apresentados os resultados para a estatística do teste de Kolmogorov-Smirnov e seu valor-p, bem como a estatística do teste de Anderson Darling e seu valor-p.

Tabela 5.3 – Testes de hipóteses.

ÍNDICE	D_n	p-valor	A^2	p-valor
IBOVESPA	0.03363	0.7752	0.4488	0.7993
S&P 500	0.03613	0.6810	0.5901	0.6573
DOW JONES	0.03558	0.6935	0.7058	0.5539

E a partir do teste de Kolmogorov-Smirnov e Anderson Darling, é possível concluir, com a análise dos p-valores obtidos, que não existem evidências estatísticas contra a hipótese de que os dados seguem as distribuições GEV transmutada aqui estimadas.

Capítulo 6

Considerações Finais

O presente trabalho teve como intuito aplicar a técnica de transmutação na distribuição generalizada de valor extremo, obtendo então uma nova distribuição denominada GEV transmutada. A transmutação é uma técnica relativamente nova (2009) e tem sido amplamente usada na literatura para introduzir assimetria em distribuições. A distribuição GEV tem grande importância em modelagem de eventos extremos com cauda pesada. Portanto, neste relatório, procura-se ampliar ainda mais a aplicabilidade da GEV, desenvolvendo a generalização da mesma por meio da transmutação quadrática.

Obtido o modelo, as características foram estudadas por meio de cálculos de medidas de probabilidades relevantes. Foram calculados os momentos, quantis, medidas de confiabilidade e estatísticas de ordem. Dentro dessas medidas, pode-se notar algumas especificidades da distribuição GEV transmutada, entre elas, a função de momentos só é capaz de fornecer momentos quando o valor do parâmetro $\gamma \in (0, 0.5]$; a função de distribuição GEV transmutada também pode ser obtida pelas estatísticas de ordem da distribuição GEV, em casos particulares.

Foram realizadas simulações da distribuição e estimações, onde, intuitivamente, pode-se concluir a partir dos resultados que existe uma relação entre os parâmetros na GEV transmutada. Essas possíveis relações não foram estudadas neste trabalho. Ficando como sugestão para continuação deste trabalho.

A GEV transmutada foi aplicada na estimação de três bancos de dados sobre índices financeiros, sendo eles, Ibovespa, S&P 500 e Dow Jones. O ajustamento da distribuição foi realizado, primeiramente, por análise visual do gráfico quantil a quantil gerado. Em seguida, foi realizado o teste de Kolmogorov-Smirnov, que conclui que a distribuição GEV

tansmutada tem um bom ajustamento para modelar os dados utilizados neste trabalho. Sendo que, as estimativas do parâmetro λ todas diferentes de 0 (zero), indicam que o modelo GEV transmutado seria melhor opção em comparação à distribuição GEV.

REFERÊNCIAS

- Aryal, G. R. (2013). Transmuted log-logistic distribution. *Journal of Statistics Applications & Probability*, 2(1), 11-20.
- Aryal, G. R., & Tsokos, C. P. (2009). On the transmuted extreme value distribution with application. *Nonlinear Analysis: Theory, Methods & Applications*, 71(12), 1401-1407.
- Aryal, G. R., & Tsokos, C. P. (2011). Transmuted Weibull Distribution: A Generalization of the Weibull Probability Distribution. *European Journal of Pure and Applied Mathematics*, 4(2), 89-102.
- Fisher, R. A. and Tippett (1928), L. H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, vol 24, pp. 180-290.
- Jenkinson, A. F (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Quarterly Journal of the Royal Meteorological Society*, London, v.81, n.551.501.45, p.158-171.
- Kotz, Samuel and Nadarajah, Saralees (2000). *Extreme value distributions: Theory and Applications*. Imperial College Press, London.
- Mendes, Beatriz V. M. (2004). *Introdução à Análise de Eventos Extremos*.
- Merovci, F. (2013). Transmuted exponentiated exponential distribution. *Mathematical Sciences And Applications E-Notes*
- Merovci, F. (2014). Transmuted generalized Rayleigh distribution. *Journal of Statistics Applications and Probability*, 3(1), 9-20.
- Merovci, F., & Elbatal, I. (2013). Transmuted Lindley-geometric distribution and its applications. *Journal of Statistics Applications & Probability*, Pro. 3, N°. 1, 77-91.
- Merovci, F., & Elbatal, I. (2014). Transmuted Weibull-geometric distribution and its applications. *School of Mathematics Northwest University*, 10(1), 68-82.
- M.R. Mahmoud and R.M. Mandouh (2014). On the Transmuted Fréchet Distribution. *Journal of Applied Sciences Research*, p.5553-5561.
- Nelder, J. A. and Mead, R. (1965) A simplex algorithm for function minimization. *Computer Journal* 7, 308-313.
- O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201-202.
- Shaw, W. T. and I. R. C. Buckley (2009). *The alchemy of Propability Distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal Distribution from a Rank Transmutation Map*.